EHRLICHIA CHAFFEENSIS 28 KDA OUTER MEMBRANE PROTEIN MULTIGENE FAMILY

BACKGROUND OF THE INVENTION

Cross-Reference to Related Application

This non-provisional patent application claims benefit of provisional patent application U.S. Serial number 60/201,035, filed May 1, 2000, now abandoned.

Federal Funding Legend

This invention was produced in part using funds obtained through a grant from the National Institute of Allergy and Infectious Disease (AI31431). Consequently, the federal government has certain rights in this invention.

Field of the Invention

The present invention relates generally to the fields of microbiology, bacteriology and molecular biology. More specifically, the present invention relates to the molecular cloning and characterization of the *Ehrlichia chaffeensis* 28 kD outer membrane protein multigene family.

Description of the Related Art

Ehrlichia are small, obligatory intracellular, gram negative bacteria which reside in endosomes inside host cells. Ehrlichiae usually cause persistent infection in their natural animal hosts (Andrew and Norval, 1989, Breitschwerdt et al., 1998, Dawson et al., 1994, Dawson and Ewing, 1992, Harrus et al., 1998, Telford et al., 1996). Persistent or prolonged *Ehrlichia* infections in human hosts have also been documented (Dumler et al., 1993, Dumler and Bakken, 1996, Horowitz, et al., 1998, Roland et al. 1994). The persistent infection may be caused by the antigenic variation of the Ehrlichia *omp-2* and *p28* outer membrane protein family due to differential expression or recombination of the *msp-2* multigene

family (Palmer et al., 1994, Palmer et al., 1998) or the *p28* multigene family (Ohashi et al., 1998b, Reddy et al., 1998, Yu et al., 1999b).

The *omp-2* and *p28* are homologous gene families coding for outer membrane proteins. The *msp-2* multigene family has been identified in *A. marginale* (Palmer et al., 1994), *A. ovina* (Palmer et al., 1998), and the human granulocytotropic ehrlichiosis agent (Ijdo et al.,1998, Murphy et al.,1998). The *p28* multigene family has been found in *E. canis* group ehrlichiae including *E. canis*, *E. chaffeensis*, and *E. muris* (McBride et al., 1999a, 1999b, Ohashi et al., 1998a, 1998b, Reddy et al., 1998, Yu et al., 1999a, 1999b). The *map-1* multigene family found in *Cowdria ruminantium* is more closely related to the *p28* multigene family than to the *msp-2* multigene family, both in sequence similarity and gene organization (Sulsona et al., 1999, van Vliet et al., 1994). The *msp-2* genes are dispersed in the genome whereas the *p28/map-1* genes are located in a single locus.

To elucidate the mechanism of the host immune avoidance involving the multigene family, the critical questions that remain to be answered are how many genes are present in each multigene family and which genes are silent or active. *E. chaffeensis*

is the pathogen of an emerging disease, human monocytotropic ehrlichiosis. Recent studies have found seven homologous polymorphic *p28* genes in *E. chaffeensis* which encode proteins from 28 to 30-kDa (Ohashi et al., 1998b, Reddy et al., 1998). The seven sequenced *p28* genes were located in three loci of the *E. chaffeensis* genome. The first locus, *omp-1* contained six *p28* genes. One gene was partially sequenced (*omp1-a*) and five genes were completely sequenced (*omp-1b, -1c, -1d, -1e*, and *-1f*) (Ohashi et al., 1998b). The second locus contained a single *p28* gene (Ohashi et al., 1998b, Yu et al., 1999b). The third locus contained five *p28* genes (ORF 1 to 5). The first four open reading frames overlapped with the DNA sequences from *omp-1 c* to *omp-1f* and the fifth open reading frame overlapped with the single gene in the second locus. Therefore, the three loci could be assembled into a single locus (Reddy et al., 1998).

The prior art is deficient in the lack of the knowledge of many of the sequences of the genes in the *p28* multigene family of *E. chaffeensis*. The present invention fulfills this long-standing need and desire in the art.

SUMMARY OF THE INVENTION

The 28-kDa outer membrane proteins (P28) of *Ehrlichia chaffeensis* are encoded by a multigene family. The *p28* multigene family of *E. chaffeensis* is located in a single locus, which is easy to sequence by genome walking. The purpose of present study was to determine all the *p28* gene sequences and their transcriptional activities. There were 21 members of the *p28* multigene family located in a 23-kb DNA fragment in the *E. chaffeensis* genome. The *p28* genes were 816 to 903 nucleotides in size and were separated by intergenic spaces of 10 to 605 nucleotides. All the genes were complete and were predicted to have signal sequences. The molecular masses of the mature proteins were predicted to be 28- to 32-kDa. The amino acid sequence identity of the P28 proteins was 20-83%. Ten *p28* genes were investigated for transcriptional activity by using RT-PCR amplification of mRNA. Six of 10 tested *p28* genes were actively transcribed in cell culture grown *E. chaffeensis*. RT-PCR also indicated that each of the *p28* genes was monocistronic. These results suggest that the *p28* genes are active genes and encode polymorphic forms of the P28 proteins. In addition, the P28s were divergent among separate isolates of *E. chaffeensis*. The large

repertoire of the *p28* genes in a single ehrlichial organism and antigenic diversity of the P28 among the isolates of *E. chaffeensis* suggest that P28s may be involved in immune avoidance.

The present invention describes the molecular cloning, sequencing, characterization, and expression of the multigene locus of P28 from *Ehrlichia chaffeensis*. The present invention describes a number of newly described genes for P28 proteins including proteins having amino acid sequences selected from the group consisting of SEQ ID No. 1, SEQ ID No.2, SEQ ID No. 3, SEQ ID No. 4, SEQ ID No. 5, SEQ ID No. 6, SEQ ID No. 7, SEQ ID No. 8, SEQ ID No. 9, SEQ ID No. 10, SEQ ID No 11, SEQ ID No. 12, SEQ ID No. 13, SEQ ID No. 20 and SEQ ID No. 21. These P28 genes are contained in a single 23 kb multigene locus of *Ehrlichia chaffeensis*. The novel part of this locus are described in GenBank accession number AF230642 and GenBank accession number AF230643.

The instant invention is also directed to DNA encoding a P28 protein selected from those described above. This DNA may consist of isolated DNA that encodes a P28 protein; isolated DNA which hybridizes to DNA encoding an isolated P28 gene, and isolated DNA encoding a P28 protein which differs due to the degeneracy of the genetic code.

The instant invention is also directed to a vector comprising a P28 gene and regulatory elements necessary for expression of the DNA in a cell. This vector may be used to transfect a host cell selected from group consisting of bacterial cells, mammalian cells, plant cells and insect cells. *E. coli* is an example of a bacterial cell into which the vector may be transfected.

The instant invention is also directed to an isolated and purified *Ehrlichia chaffeensis* P28 surface protein selected from those described above including those with amino acid sequences SEQ ID No. 1, SEQ ID No.2, SEQ ID No. 3, SEQ ID No. 4, SEQ ID No. 5, SEQ ID No. 6, SEQ ID No. 7, SEQ ID No. 8, SEQ ID No. 9, SEQ ID No. 10, SEQ ID No 11, SEQ ID No. 12, SEQ ID No. 13, SEQ ID No. 20 and SEQ ID No. 21.

The instant invention also describes an antibody directed against one of these P28 proteins. This antibody may be a monoclonal antibody.

The novel P28 proteins of the instant invention may be used in a vaccine against *Ehrlichia chaffeensis.*

Other and further aspects, features, and advantages of the present invention will be apparent from the following

description of the presently preferred embodiments of the invention given for the purpose of disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

So that the matter in which the above-recited features, advantages and objects of the invention, as well as others which will become clear, are attained and can be understood in detail, more particular descriptions of the invention briefly summarized above may be had by reference to certain embodiments thereof which are illustrated in the appended drawings. These drawings form a part of the specification. It is to be noted, however, that the appended drawings illustrate preferred embodiments of the invention and therefore are not to be considered limiting in their scope.

Figure 1 shows the scheme of sequencing the *p28* gene locus by genome walking and the organization of the *p28* genes. Three loci of p28 genes previously sequenced were aligned and assembled into a single contiguous sequence. Initial primers (arrow heads) were designed near the 5' and 3' ends of the contiguous

sequence to walk the genome. The block arrows represented the positions and the directions of the *p28* genes. The scale indicated the nucleotides in kilobases.

Figure 2 shows a clustal alignment of the amino acid sequences of the *E. chaffeensis* Arkansas strain P28s (1-21). P28-1 was used as consensus sequence. Dots represented residues identical to those of the consensus sequence. Gaps represented by dash lines were introduced for optimal alignment of the DNA sequences. The hypervariable regions were underlined.

Figure 3 shows the phylogenetic relationships of the P28s (1-21). The number on the branch indicated the bootstrap values.

Figure 4 shows Southern blotting. Two bands of 17.6 and 5.3 kb were detected by a *p28* gene probe on *Cla* I restriction endonuclease digested *E. chaffeensis* genomic DNA (lane E). M: molecular weight marker.

Figure 5 shows RT-PCR amplification of the mRNA of *E. chaffeensis p28* genes (RT-PCR). In the PCR controls, reverse transcriptase was omitted. The numbers of each lane indicated the *p28* genes. M represents a molecular weight marker.

DETAILED DESCRIPTION OF THE INVENTION

The following abbreviations may be used herein: BCIP/NBT-5-bromo-4-chloro-3-indolylphosphate/ nitrobluetetrazolium substrate; ATP - adenosine triphosphate; DNA - deoxyribonucleic acid; E - *Ehrlichia*; kDa - kilodalton; mRNA - messenger ribonucleic acid; ORF - open reading frame; P28 - 28-kDa outer membrane proteins; PCR - polymerase chain reaction; RT-PCR - reverse transcriptase-polymerase chain reaction.

In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Maniatis, Fritsch & Sambrook, "Molecular Cloning: A Laboratory Manual (1982); "DNA Cloning: A Practical Approach," Volumes I and II (D.N. Glover ed. 1985); "Oligonucleotide Synthesis" (M.J. Gait ed. 1984); "Nucleic Acid Hybridization" [B. D. Hames & S.J. Higgins eds. (1985)]; "Transcription and Translation" [B. D. Hames & S.J. Higgins eds. (1984)]; "Animal Cell Culture" [R. I. Freshney, ed. (1986)];

"Immobilized Cells And Enzymes" [IRL Press, (1986)]; B. Perbal, "A Practical Guide To Molecular Cloning" (1984).

Therefore, if appearing herein, the following terms shall have the definitions set out below.

A "replicon" is any genetic element (e.g., plasmid, chromosome, virus) that functions as an autonomous unit of DNA replication *in vivo*; i.e., capable of replication under its own control.

A "vector" is a replicon, such as plasmid, phage or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment.

A "DNA molecule" refers to the polymeric form of deoxyribonucleotides (adenine, guanine, thymine, or cytosine) in its either single stranded form, or a double-stranded helix. This term refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear DNA molecules (e.g., restriction fragments), viruses, plasmids, and chromosomes. In discussing the structure herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (i.e., the strand having a sequence homologous to the mRNA).

A DNA "coding sequence" is a double-stranded DNA sequence that is transcribed and translated into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from eukaryotic (e.g., mammalian) DNA, and even synthetic DNA sequences. A polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

Transcriptional and translational control sequences are DNA regulatory sequences, such as promoters, enhancers, polyadenylation signals, terminators, and the like, that provide for the expression of a coding sequence in a host cell.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of

bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site, as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters often, but not always, contain "TATA" boxes and "CAT" boxes. Prokaryotic promoters contain Shine-Dalgarno sequences in addition to the -10 and -35 consensus sequences.

An "expression control sequence" is a DNA sequence that controls and regulates the transcription and translation of another DNA sequence. A coding sequence is "under the control" of transcriptional and translational control sequences in a cell when RNA polymerase transcribes the coding sequence into mRNA, which is then translated into the protein encoded by the coding sequence.

A "signal sequence" can be included near the coding sequence. This sequence encodes a signal peptide, N-terminal to the polypeptide, that communicates to the host cell to direct the polypeptide to the cell surface or secrete the polypeptide into the media, and this signal peptide is clipped off by the host cell before the protein leaves the cell. Signal sequences can be found

associated with a variety of proteins native to prokaryotes and eukaryotes.

The term "oligonucleotide", as used herein in referring to the probe of the present invention, is defined as a molecule comprised of two or more ribonucleotides, preferably more than three. Its exact size will depend upon many factors which, in turn, depend upon the ultimate function and use of the oligonucleotide.

The term "primer" as used herein refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically. A "primer" is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product, which is complementary to a nucleic acid strand, is induced (i.e., in the presence of nucleotides and an inducing agent such as a DNA polymerase and at a suitable temperature and pH). The primer may be either single-stranded or double-stranded and must be sufficiently long to prime the synthesis of the desired extension product in the presence of the inducing agent. The exact length of the primer will depend upon many factors, including temperature, source of primer and use the method. For example, for diagnostic applications, depending on the complexity of the target sequence,

the oligonucleotide primer typically contains 15-25 or more nucleotides, although it may contain fewer nucleotides.

The primers herein are selected to be "substantially" complementary to different strands of a particular target DNA sequence. This means that the primers must be sufficiently complementary to hybridize with their respective strands. Therefore, the primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being complementary to the strand. Alternatively, non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence or hybridize therewith and thereby form the template for the synthesis of the extension product.

A cell has been "transformed" by exogenous or heterologous DNA when such DNA has been introduced inside the cell. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With

respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

Two DNA sequences are "substantially homologous" when at least about 75% (preferably at least about 80%, and most preferably at least about 90% or 95%) of the nucleotides match over the defined length of the DNA sequences. Sequences that are substantially homologous can be identified by comparing the sequences using standard software available in sequence data banks, or in a Southern hybridization experiment under, for example, stringent conditions as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., Maniatis et al., *supra*; DNA Cloning, Vols. I & II, *supra*; Nucleic Acid Hybridization, *supra*.

A "heterologous' region of the DNA construct is an identifiable segment of DNA within a larger DNA molecule that is not found in association with the larger molecule in nature. Thus, when the heterologous region encodes a mammalian gene, the gene will usually be flanked by DNA that does not flank the mammalian genomic DNA in the genome of the source organism. In another example, coding sequence is a construct where the coding sequence itself is not found in nature (e.g., a cDNA where the genomic coding sequence contains introns, or synthetic sequences having codons different than the native gene). Allelic variations or naturally occurring mutational events do not give rise to a heterologous region of DNA as defined herein.

The labels most commonly employed for these studies are radioactive elements, enzymes, chemicals which fluoresce when exposed to ultraviolet light, and others. A number of fluorescent materials are known and can be utilized as labels. These include, for example, fluorescein, rhodamine, auramine, Texas Red, AMCA blue and Lucifer Yellow. A particular detecting material is anti-rabbit antibody prepared in goats and conjugated with fluorescein through an isothiocyanate.

Proteins can also be labeled with a radioactive element or with an enzyme. The radioactive label can be detected by any of the currently available counting procedures. The preferred isotope may be selected from ^{3}H, ^{14}C, ^{32}P, ^{35}S, ^{36}Cl, ^{51}Cr, ^{57}Co, ^{58}Co, ^{59}Fe, ^{90}Y, ^{125}I, ^{131}I, and ^{186}Re.

Enzyme labels are likewise useful, and can be detected by any of the presently utilized colorimetric, spectrophotometric, fluorospectrophotometric, amperometric or gasometric techniques. The enzyme is conjugated to the selected particle by reaction with bridging molecules such as carbodiimides, diisocyanates, glutaraldehyde and the like. Many enzymes which can be used in these procedures are known and can be utilized. The preferred are peroxidase, β-glucuronidase, β-D-glucosidase, β-D-galactosidase, urease, glucose oxidase plus peroxidase and alkaline phosphatase. U.S. Patent Nos. 3,654,090, 3,850,752, and 4,016,043 are referred to by way of example for their disclosure of alternate labeling material and methods.

As used herein, the term "host" is meant to include not only prokaryotes but also eukaryotes such as yeast, plant and animal cells. A recombinant DNA molecule or gene which encodes a 28-kDa immunoreactive protein of *Ehrlichia chaffeensis* of the

present invention can be used to transform a host using any of the techniques commonly known to those of ordinary skill in the art. Especially preferred is the use of a vector containing coding sequences for a gene encoding a 28-kDa immunoreactive protein of *Ehrlichia chaffeensis* of the present invention for purposes of prokaryote transformation.

Prokaryotic hosts may include *E. coli*, *S. tymphimurium*, *Serratia marcescens* and *Bacillus subtilis*. Eukaryotic hosts include yeasts such as *Pichia pastoris*, mammalian cells and insect cells.

In general, expression vectors containing promoter sequences which facilitate the efficient transcription of the inserted DNA fragment are used in connection with the host. The expression vector typically contains an origin of replication, promoter(s), terminator(s), as well as specific genes that are capable of providing phenotypic selection in transformed cells. The transformed hosts can be fermented and cultured according to means known in the art to achieve optimal cell growth.

By "high stringency" is meant DNA hybridization and wash conditions characterized by high temperature and low salt concentration, e.g., wash conditions of 65°C at a salt concentration of approximately 0.1 x SSC, or the functional equivalent thereof.

For example, high stringency conditions may include hybridization at about 42°C in the presence of about 50% formamide; a first wash at about 65°C with about 2 x SSC containing 1% SDS; followed by a second wash at about 65°C with about 0.1 x SSC.

By "substantially pure DNA" is meant DNA that is not part of a milieu in which the DNA naturally occurs, by virtue of separation (partial or total purification) of some or all of the molecules of that milieu, or by virtue of alteration of sequences that flank the claimed DNA. The term therefore includes, for example, a recombinant DNA which is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote; or which exists as a separate molecule (e.g., a cDNA or a genomic or cDNA fragment produced by polymerase chain reaction (PCR) or restriction endonuclease digestion) independent of other sequences. It also includes a recombinant DNA that is part of a hybrid gene encoding additional polypeptide sequence, e.g., a fusion protein.

The identity between two sequences is a direct function of the number of matching or identical positions. When a subunit position in both of the two sequences is occupied by the same monomeric subunit, e.g., if a given position is occupied by an

adenine in each of two DNA molecules, then they are identical at that position. For example, if 7 positions in a sequence 10 nucleotides in length are identical to the corresponding positions in a second 10-nucleotide sequence, then the two sequences have 70% sequence identity. The length of comparison sequences will generally be at least 50 nucleotides, preferably at least 60 nucleotides, more preferably at least 75 nucleotides, and most preferably 100 nucleotides. Sequence identity is typically measured using sequence analysis software (e.g., Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705).

A "vector" may be defined as a replicable nucleic acid construct, e.g., a plasmid or viral nucleic acid. Vectors may be used to amplify and/or express nucleic acid encoding a 28-kDa immunoreactive protein of *Ehrlichia chaffeensis*. An expression vector is a replicable construct in which a nucleic acid sequence encoding a polypeptide is operably linked to suitable control sequences capable of effecting expression of the polypeptide in a cell. The need for such control sequences will vary depending upon the cell selected and the transformation method chosen. Generally,

control sequences include a transcriptional promoter and/or enhancer, suitable mRNA ribosomal binding sites, and sequences which control the termination of transcription and translation. Methods, which are well known to those skilled in the art, can be used to construct expression vectors containing appropriate transcriptional and translational control signals. See for example, the techniques described in Sambrook et al., 1989, *Molecular Cloning: A Laboratory Manual* (2nd Ed.), Cold Spring Harbor Press, N.Y. A gene and its transcription control sequences are defined as being "operably linked" if the transcription control sequences effectively control the transcription of the gene. Vectors of the invention include, but are not limited to, plasmid vectors and viral vectors. Preferred viral vectors of the invention are those derived from retroviruses, adenovirus, adeno-associated virus, SV40 virus, or herpes viruses.

By a "substantially pure protein" is meant a protein that has been separated from at least some of those components that naturally accompany it. Typically, the protein is substantially pure when it is at least 60%, by weight, free from the proteins and other naturally occurring organic molecules with which it is naturally associated in vivo. Preferably, the purity of the preparation is at

least 75%, more preferably at least 90%, and most preferably at least 99%, by weight. A protein is substantially free of naturally associated components when it is separated from at least some of those contaminants that accompany it in its natural state. Thus, a protein that is chemically synthesized or produced in a cellular system different from the cell from which it naturally originates will be, by definition, substantially free from its naturally associated components. Accordingly, substantially pure proteins include eukaryotic proteins synthesized in *E. coli*, other prokaryotes, or any other organism in which they do not naturally occur.

The phrase "pharmaceutically acceptable" refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a human. The preparation of an aqueous composition that contains a protein as an active ingredient is well understood in the art. Typically, such compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection can also be prepared. The preparation can also be emulsified.

A protein may be formulated into a composition in a neutral or salt form. Pharmaceutically acceptable salts, include the

acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like.

Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms such as injectable solutions.

For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal administration. In this connection, sterile aqueous media that can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage could be dissolved in 1

ml of isotonic NaCl solution and either added to 1000mL of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject.

As is well known in the art, a given polypeptide may vary in its immunogenicity. It is often necessary therefore to couple the immunogen (e.g., a polypeptide of the present invention) with a carrier. Exemplary and preferred carriers are keyhole limpet hemocyanin (KLH) and human serum albumin. Other carriers may include a variety of lymphokines and adjuvants such as IL2, IL4, IL8 and others.

Means for conjugating a polypeptide to a carrier protein are well known in the art and include glutaraldehyde, m-maleimidobenzoyl-N-hydroxysuccinimide ester, carbo-diimide and bis-biazotized benzidine. It is also understood that the peptide may be conjugated to a protein by genetic engineering techniques that are well known in the art.

As is also well known in the art, immunogenicity to a particular immunogen can be enhanced by the use of non-specific stimulators of the immune response known as adjuvants. Exemplary and preferred adjuvants include complete BCG, Detox, RIBI (Immunochem Research Inc.), ISCOMS and aluminum hydroxide adjuvant (Superphos, Biosector).

As used herein the term "complement" is used to define the strand of nucleic acid which will hybridize to the first nucleic acid sequence to form a double stranded molecule under stringent conditions. Stringent conditions are those that allow hybridization between two nucleic acid sequences with a high degree of homology, but precludes hybridization of random sequences. For example, hybridization at low temperature and/or high ionic strength is termed low stringency and hybridization at high temperature and/or low ionic strength is termed high stringency. The temperature and ionic strength of a desired stringency are understood to be applicable to particular probe lengths, to the length and base content of the sequences and to the presence of formamide in the hybridization mixture.

As used herein, the term "engineered" or "recombinant" cell is intended to refer to a cell into which a recombinant gene,

such as a gene encoding an *Ehrlichia chaffeensis* antigen has been introduced. Therefore, engineered cells are distinguishable from naturally occurring cells that do not contain a recombinantly introduced gene. Engineered cells are thus cells having a gene or genes introduced through the hand of man. Recombinantly introduced genes will either be in the form of a cDNA gene, a copy of a genomic gene, or will include genes positioned adjacent to a promoter not naturally associated with the particular introduced gene. In addition, the recombinant gene may be integrated into the host genome, or it may be contained in a vector, or in a bacterial genome transfected into the host cell.

The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion.

EXAMPLE 1

Ehrlichia spp

Ehrlichia chaffeensis (Arkansas strain) was obtained from Jacqueline Dawson (Centers for Disease Control and

Prevention, Atlanta, GA). Ehrlichiae were cultivated in DH82 cells, a canine macrophage-like cell line. DH82 cells were harvested with a cell scraper when 100% of cells were infected with ehrlichiae. The cells were centrifuged at 17,400 X g for 20 min. The pellets were disrupted twice with a Braun-Sonic 2000 sonicator at 40 W for 30 sec on ice. *Ehrlichia* were then purified by using 30% Percoll gradient centrifugation (Weiss et al, 1989).

EXAMPLE 2

PCR amplification of the *p28* multigene locus

Ehrlichia chaffeensis genomic DNA was prepared by using an IsoQuick Nucleic Acid Extraction Kit (ORCA Research Inc., Bothell, WA) according to the instructions of the manufacturer. The unknown sequences of the *p28* multigene locus were amplified by PCR using the Universal GenomeWalker Kit (Clontech Laboratories, Inc., Palo Alto, CA). Briefly, the *E. chaffeensis* genomic DNA was digested respectively with *Dra* I, *EcoR* V, *Pvu* II, *Sca* I, and *Stu* I. The enzymes were chosen because they generated blunt ended DNA fragments to ligate with the blunt-end of the adapter. The digested

E. chaffeensis genomic DNA fragments were ligated with a GenomeWalker Adapter, which had one blunt end and one end with 5' overhang. The ligation mixture of the adapter and *E. chaffeensis* genomic DNA fragments was used as template for PCR. Initially, the *p28* gene-specific primer amplified the known DNA sequence and extended into the unknown adjacent genomic DNA and the adapter 5'overhang, which is complementary to the adapter primer. In the subsequent PCR cycles, the target DNA sequences were amplified with both the *p28* gene-specific primer and the adapter primer.

EXAMPLE 3

DNA sequencing

The PCR products were purified by using a QIAquick PCR Purification Kit (QIAGEN Inc., Santa Clarita, CA) and were sequenced directly using PCR primers when a single clear band was observed on the ethidium-bromide stained agarose gel. If multiple bands appeared, the DNA band of interest was excised from the gel, and the DNA was extracted from the gel using the Gel Extraction Kit (QIAGEN Inc., Santa Clarita, CA). The gel-purified DNA was cloned

into the Topo TA cloning vector (Invitrogen, Inc., Carlsbad, CA) according to the instructions of the manufacturer. A High Pure Plasmid Isolation Kit (Boehringer Mannheim Corp., Indianapolis, IN) was used to purify the plasmids. An ABI Prism 377 DNA Sequencer (Perkin-Elmer Applied Biosystems, Foster City, CA) was used to sequence the DNA in the Protein Chemistry Laboratory of the University of Texas Medical Branch.

EXAMPLE 4

Gene analysis

DNA sequences and deduced amino acid sequences were analyzed using DNASTAR software (DNASTAR, Inc., Madison, WI). The signal sequence of the deduced protein was analyzed by using the PSORT program, which predicts the presence of signal sequences (McGeoch, 1985, Von Heijne, 1986) and detects potential transmembrane domains (Klein, 1985). Phylogenetic analysis was performed by the maximum parsimony method of the PAUP 4.0 software (Sunderland Massachusetts: Sinauer Associates, 1998).

Bootstrap values for the consensus tree were based on analysis of 1000 replicates.

EXAMPLE 5

DNA sequence accession numbers

The DNA sequences of the *E. chaffeensis p28* genes were assigned GenBank accession numbers: AF230642 for the DNA locus of the *p28-1* to *p28-13* and AF230643 for the DNA locus of *p28-20* and *p28-21*.

EXAMPLE 6

Reverse transcriptase PCR (RT-PCR)

Total RNA of *E. chaffeensis*-infected DH82 cells was isolated using RNeasy Total RNA Isolation Kit (Qiagen Inc., Santa Clarita, Ca). The *p28* gene mRNA (0.5 μg total RNA) was amplified using a Titan One Tube RT-PCR System (Roche Molecular Biochemicals, Indianapolis, IN) according to the manufacturer's

instructions. Gene-specific primer pairs used in the RT-PCR reaction were listed in Table 1. A negative control that included all reagents except reverse transcriptase was included to confirm that genomic DNA was not present in the total RNA preparation. The thermal cycling profile consisted of reverse transcription at 50° C for 30 min, amplification for 30 cycles at 94° C for 2 min, 50° C for 1 min, and 68° C for 1 min, and an elongation step at 68° C for 7 min.

TABLE 1

Gene-specific primers for RT-PCR

Gene	Sequences of forward (f) and reverse (r) primers	Product length (bp)
p28-10	(f)ACG TGA TAT GGA AAG CAA CAA GT (SEQ ID No. 22) (r)GCG CCG AAA TAT CCA ACA (SEQ ID No. 23)	384
p28-11	(f)GGT CAA ACT TGC CCT AAA CAC A (SEQ ID No. 24) (r)ACT TCA CCA CCA AAA TAC CCA ATA (SEQ ID No. 25)	406
p28-12	(f)CTG CTG GCA TTA GTT ACC C (SEQ ID No. 26) (r)CAT AGC AGC CAT TGA CC (SEQ ID No. 27)	334
p28-13	(f) ATT GAT TGC CTA TTA CTT GAT GGT (SEQ ID No. 28) (r)AAT GGG GCT GTT GGT TAC TC (SEQ ID No. 29)	333
p28-14	(f)TGA AGA CGC AAT AGC AGA TAA GA (SEQ ID No. 30) (r)TAG CGC AGA TGT GGT TTG AG (SEQ ID No. 31)	269
p28-15	(f) ACT GTC GCG TTG TAT GGT TTG (SEQ ID No. 32) (r)ATT AGT GCT GCT TGC TTT ACG A (SEQ ID No. 33)	371
p28-17	(f) TGC AAG GTG ACA ATA TTA GTG GTA (SEQ ID No. 34) (r) GTA TTC CGC TGT TGT CTT GTT G (SEQ ID No. 35)	367
p28-18	(f)ACA TTT TGG CGT ATT CTC TGC (SEQ ID No. 36) (r)TAG CTT TCC CCC ACT GTT ATG (SEQ ID No. 37)	312
p28-20	(f)AAC TTA TGG CTT TCT CCT CCT TTC (SEQ ID No. 38) (r)TTG CCT GAT AAT TCT TTT TCT GAT (SEQ ID No. 39)	340
p28-21	(f)ACC AAC TTC CCA ACC AAA ATA ATC (SEQ ID No. 40) (r) CTG AAG GAG GAG AAA GCC ATA AGT (SEQ ID No. 41)	421

EXAMPLE 7

Southern blotting

The DNA sequences of the *p28* multigene locus were analyzed for the presence of restriction sites using a Mapdraw program (DNASTAR, Inc., Madison, WI). *Ehrlichia chaffeensis* genomic DNA was digested by restriction endonuclease *Cla* I. The DNA was separated using a 0.8% agarose gel. DNA was blotted onto nylon membranes by capillary transfer. The probe was DNA-amplified from the *p28* multigene locus by using PCR and was labeled with digoxigenin-11-dUTP using a DIG DNA Labeling Kit (Roche Molecular Biochemicals, Indianapolis, IN). The probe corresponded to the nucleotides from 8900 to 10620 of the locus, which included the 3' end of *p28-7*, the entire gene of *p28-8*, the 5' end of *p28-9*, and the intergenic sequences between the three genes. DNA hybridization was performed at 42°C overnight in the Eazy Hybridization Buffer (Roche Molecular Biochemicals, Indianapolis, IN). The DNA probes were detected using the colorimetric reagent (BCIP/NBT) following the instructions of the manufacturer (Roche Molecular Biochemicals, Indianapolis, IN).

EXAMPLE 8

PCR amplification of the *p28* multigene locus

The sequences of three *p28* gene loci were obtained from GenBank (accessions: AF021338, AF062761, and AF068234) (Ohashi et al., 1998b, Reddy et al., 1998, Yu et al., 1999b) and were assembled into a single contiguous DNA sequence which contained seven *p28* genes with the first one incomplete. Gene-specific primers to the partial gene (primer 1a-r1 and primer1a-r2) and the DNA sequence downstream of the last *p28* gene (primers 28f1 and 28f2) were designed from the contiguous sequence for the initial extension of the *p28* gene locus of *E. chaffeensis*.

The scheme of PCR-amplification of the *p28* multigene locus is illustrated in Fig. 1, and the sequences of the gene specific primers were listed in Table 2. A 1.6-kb DNA fragment was amplified initially from the 5' end of the locus from a *Stu* I-restriction genomic library by nested PCR using primer 1a-r2. The PCR products were sequenced directly, and a new primer (28r3) was designed from the sequence to further extend the 5' end sequence of the locus. A 4.5-kb DNA fragment (pvu4.5) was amplified from a *Pvu* II-restriction genomic library by using primer 28r3. The 5' end

of the DNA locus was further extended with six additional primer walks by using primers: pvur32, 28r12, 28stur, 28r14, and 28r15. Each primer was designed from the DNA sequences from the preceding PCR product. The 3' end of the locus was initially extended for 1.5-kb by nested PCR using primers 28f1 and 28f2. The 1.5-kb DNA fragment was directly sequenced and used to design a new primer (28f3) to further walk the 3' end of the locus.

A 2.8-kb DNA fragment (stu2.8) was amplified from a *Stu* I-restriction genomic library by using primer 28f3. The pvu4.5, pvu1.8, and stu2.8 DNA fragments were gel-purified and cloned into the Topo TA PCR cloning vector. The DNA in the Topo TA vector was sequenced initially using the M13 reverse and M13 forward primers and extended by primer walking. The sequence on the 5' end of stu2.8 was not readable following M13 forward and reverse primers, possibly due to the secondary structure. Thus, the recombinant Topo TA plasmid containing the stu2.8 DNA was digested with the restriction enzyme *Kpn* I. A 700-bp fragment of DNA was deleted from the 5' end of the stu2.8 DNA. The plasmid was ligated again, and the insert was sequenced using M13 reverse and M13 forward primers. The rest of PCR products were sequenced directly.

TABLE 2

Primers for genome walking the *E. chaffeensis p28* multigene locus

Name	Sequences	Product length (kb)
1a-r1[a]	ACC AAA GTA TGC AAT GTC AAG TG (SEQ ID No.42)	
1a-r2	CTG CAG ATG TGA CTT TAG GAG ATT C (SEQ ID No.43)	1.6
28r3	TGT ATA TCT TCC AGG GTC TTT GA (SEQ ID No.44)	4.5
pvur32	GAC CAT TCT ACC TCA ACC (SEQ ID No.45)	1.8
28r10	ATA TCC AAT TGC TCC ACT GAA A (SEQ ID No.46)	1.5
28r12	CTT GAA ATG TAA CAG TAT ATG GAC CTT GAA (SEQ ID No.47)	2.2
28stur	TGT CCT TTT TAA GCC CAA CT (SEQ ID No.48)	1.5
28r14	TTC TGC AGA TTG ATG TGG ATG TTT (SEQ ID No.49)	4.7
28r15	TGC AGA TTG ATG TGG ATG TTT (SEQ ID No.50)	1.1
28f1[b]	GTA AAA CAC AAG CCA CCA GTC T (SEQ ID No.51)	
28f2	GGG CAT ATA CCT ACA CCA AAC ACC (SEQ ID No.52)	1.5
28f3	TAA GAG GAT TGG GTA AGG ATA (SEQ ID No.53)	2.8

a: 1a-r1 was outside primer for1a-r2; b: 28f1 was outside primer for 28f2.

EXAMPLE 9

p28 gene family consists of 21 homologous but distinct genes

The sequences of the DNA fragments were assembled together by using the Seqman program (DNASTAR, Inc., Madison, WI) into a 23-kb segment of DNA. There were 21 homologous *p28* genes in the DNA locus. The genes were designated as *p28-1* to *p28-21* according to their positions from the 5' end to the 3' end of the locus (Fig. 1). Most of the genes were tandemly arranged in one direction in the locus, and the last two genes (*p28-20* and *p28-21*) were in the complementary strand. The sizes of the genes ranged from 816 bp to 903 bp while length of the non-coding sequences between the neighboring genes varied from 10 to 605-bp. The intergenic spaces between *p28*-1 and *p28*-2 and between *p28-6* and *p28-7* encoded a 150 amino acid protein and a 195 amino acid protein, respectively, and the two proteins had no sequence similarity to any known proteins. On the 5' end of the locus, there is a 1347 nucleotide open reading frame, which was similar to *clpX* gene, a class-III heat-shock gene encoding an ATP-dependent protease.

All the P28s were predicted to have a signal sequence. The signal sequences of P28-1, P28-7, and P28-8 were predicted to be uncleavable. The signal sequences of the rest of the P28s were predicted to be cleavable, and the proteins were predicted to be cleaved from positions varying from position 19 to position 30. The predicted molecular sizes of the mature P28s were from 25.8-kDa to 32.1-kDa. The C-termini and the middle of the proteins were most conserved. There were 4 hypervariable regions in the amino acid sequences of the P28 proteins (Fig. 2). The first hypervariable region was immediately after the signal sequence. No proteins had identical sequences in the hypervariable regions (Fig. 2).

EXAMPLE 10

Phylogenetic relationships of the P28s

The amino acid sequence identity of the P28s varied from 20% to 83% (Fig. 3). In general, the proteins derived from adjacent genes had higher identities. The P28s having the highest amino acid sequence identities were from P28-16 to P28-19, which were 68.3 to 82.7% identical to each other. The next group with

high sequence identity was from P28-7 to P28-13, which were 47.6 to 66.9% identical to each other. The sequence identity among the rest of the *E. chaffeensis* P28s were from 19.7 to 45.6%.

The amino acid sequences of the P28s of *E. chaffeensis* were highly homologous to the P28 protein families of *E. canis* and *E. muris* (McBride et al., 1999a, 1999b, Reddy et al., 1998, Yu et al., 1999a) and the MAP-1 protein family of *C. ruminantium* (Van Vliet et al.,1994, Sulsona et al., 1999). P28-17 of *E. chaffeensis* was the most conserved protein among the *Ehrlichia* species. The amino acid sequence of the *E. chaffeensis* P28-17 was 58% to 60% identical to the P28s of *E. canis* and 78% to 81% identical to the P28s of *E. muris*. The P28s of *E. chaffeensis* also have significant similarity to the MSP-4 protein (Oberle and Barbet, 1993), and the MSP-2 protein families of *A. marginale* (Palmer et al., 1994) and the MSP-2 of the human granulocytotropic ehrlichiosis agent (Ijdo et al., 1998, Murphy et al., 1998).

EXAMPLE 11

p28 genes located in a single locus

Southern blotting was performed to detect whether all the *p28* genes were located on a single locus and whether the whole locus has been sequenced. *Cla* I restriction endonuclease was predicted to digest the *p28* gene locus at three sites generating 5268 bp and 17550 bp DNA fragments. Southern blot using a *p28* gene probe demonstrated a strong band of 17.6-kb and a weak band of 5.3-kb in the *Cla* I-digested *E. chaffeensis* genomic DNA (Fig. 4). This result indicated that all the *p28* genes were located on two *Cla* I DNA fragments and that all the *p28* genes had been sequenced. Sequencing a segment of 2.3 kb DNA upstream of the first *p28* gene and a segment of 2 kb downstream of the last *p28* gene did not reveal any additional *p28* genes.

EXAMPLE 12

Transcriptional activity of the *p28* multigene family

The transcriptional activity was evaluated by RT-PCR for 10 *p28* genes including *p28-10*, *p28-11*, *p28-12*, *p28-13*, *p28-14*,

p28-15, *p28-17*, *p28-18*, *p28-20*, and *p28-21* (Fig. 5). These genes were selected for transcriptional analysis because they represented genes tightly clustered together (*p28-10* to *p28-13*), genes with larger intergenic spaces (*p28-14* to *p28-18*), or genes in the complementary strand (*p28-20* and *p28-21*). To ensure the specificity of RT-PCR, each primer pair was designed to be specific for a single *p28* gene only. DNA bands of expected size were observed in ethidium-bromide stained agarose gels of the RT-PCR products for the following genes: *p28-10*, *p28-11*, *p28-12*, *p28-15*, *p28-18*, and *p28-20*. No DNA band was detected in ethidium-bromide stained agarose gels of RT-PCR products of the following genes: *p28-13*, *p28-14*, *p28-17*, and *p28-21*. The rest of the *p28* genes were not investigated for their transcription. In the controls, no DNA was amplified from any genes by PCR reactions from which reverse transcriptase was omitted. All the primer pairs produced products of the expected size when using *E. chaffeensis* genomic DNA as template (data not shown).

EXAMPLE 13

p28 genes were monocistronic

Monocistronic mRNA represents a single gene and polycistronic mRNA codes for several proteins. Two adjacent *p28* genes might be polycistronically transcribed if both genes yield RT-PCR products. Two adjacent genes were monocistronically transcribed if one gene yielded a RT-PCR product and the other yielded no RT-PCR product. From Fig. 5, it was deduced that the following pairs of genes were not polycistronically transcribed: *p28-12* and *p28-13*, *p28-14* and *p28-15*, *p28-17* and *p28-18*, and *p28-20* and *p28-21*. The detection of *p28-10* to *p28-12* by RT-PCR indicated they might have been transcribed polycistronically. However, a RT-PCR experiments using the *p28-10* gene forward primer and the *p28-11* gene reverse primer failed to produce any PCR product. Furthermore, amplification with the *p28-11* gene forward primer and the *p28-12* gene reverse primer to amplify *p28-11* and *p28-12* as a single DNA fragment failed to yield product. However, both pairs of primers amplified the corresponding DNA segments. This data indicated that these genes were monocistronically transcribed.

EXAMPLE 14

The P28s were divergent among the *E. chaffeensis* isolates

A *p28* gene corresponding to *p28-19* of Arkansas strain was sequenced in four additional *E. chaffeensis* isolates made previously (Yu et al., 1999b). Clustal alignment indicated that none of the P28 genes of the Arkansas strain had identical amino acid sequence with the single sequenced P28 of the four *E. chaffeensis* isolates. The sequenced P28's from all four isolates were most similar (85-86%) to the P28-19 protein of Arkansas strain. Thus, they were analogs of P28-19 of Arkansas strain.

Discussion

The complete sequence of an entire locus of *p28* genes is reported herein for the first time. Complete sequencing of the *p28* multigene locus in *E. chaffeensis* in this study will contribute to the investigation of the origin of the multigene family and the function of the multigenes. Gene families are thought to have arisen by duplication of an original ancestral gene, with different members of the family then diverging as a consequence of mutations during evolution. The most conserved *p28* gene among the species of

Ehrlichia should be the ancestral gene. *E. chaffeensis* *p28-15* to *p28-19* are the genes most similar to the *p28* of *E. canis* and *E. muris*. Therefore, the *p28* genes might have arisen from one of the *p28-15* to *p28-19* genes. The wide presence of the *p28/msp-2* multigenes in the *Ehrlichia*, *Anaplasma*, and *Cowdria* indicate that these organisms are phylogenetically related. The significant sequence identity between the *p28* multigene family and the *msp-2* multigene family indicates that the two gene families originated from a common ancestor gene.

p28 genes corresponding to the *p28-14* to *p28-19* were sequenced previously and designated as *omp-1b* to *omp-1f* and *p28* by Ohashi et al. (1998b) and ORF-1 to ORF-5 by Reddy et al(1998). An alphabetic letter or a number assigned to each gene attempted to indicate the order and position of the genes in the locus. Neither previously assigned letters nor the numbers truly represent the position of the genes in the locus as revealed when it was sequenced completely. Thus, the genes were renamed to best represent the order of the genes in the complete locus. P28 was used as the name of the protein because it accurately describes the molecular mass of an immunodominant protein which was determined before its gene was sequenced (Chen et al., 1994, Yu et al., 1993) and also because

the *p28* was used to describe its gene name when the first *p28* gene was cloned and sequenced (Ohashi et al.,1998b).

Six *p28* genes were expressed in cell culture under the particular conditions of the investigation among the 10 genes studied. The genes for which transcription were not detected by RT-PCR are possibly not silent genes either since all the genes were complete genes, i.e., no truncated form of the *p28* genes was found. They may be expressed under other conditions. These results were consistent with previous data, which detected multiple bands from 22-29 kDa with a monoclonal antibody (Yu et al., 1993, 1999b). In contrast, a previous study detected only a single *p28* gene transcribed in cell culture (Reddy et al.,1998). PCR primer specificity may have contributed to the failure of detection the transcription of multiple genes in the previous study. With the limitation of knowledge of the DNA sequences at that time, although primers were designed to attempt to amplify as many *p28* genes as possible, the primer pair (R72 and R74) from the previous study was perfectly matched to only three of the 21 *p28* genes (*p28*-16, -17, and -19). The previous study demonstrated that *p28-19* (orf-5) was transcriptionally active and *p28-16* and *p28-17* were inactive transcriptionally. In the results herein. *p28-17* was also

transcriptionally inactive. The transcriptional activity of *p28-16* and *p28-19* was not analyzed. It was possible to detect transcriptional activity in more *p28 genes* herein because specific primers were used for each *p28* gene.

The natural cycles of *Ehrlichia* involve a tick vector and mammalian hosts. Mammals are infected with *Ehrlichia* by the bite of infected ticks, and non-infected ticks acquire *Ehrlichia* by a blood meal from infected animals. *Ehrlichia* are not transovarially transmitted from one generation of ticks to the next (Rikihisa, 1991). Therefore, the mammalian hosts are essential for the maintenance of *Ehrlichia* in nature. Carrier animals serve as the reservoirs for *Ehrlichia* organisms (Swift and Thomas, 1983, Zaugg, et al., 1986). The persistent infection and carrier status indicate that *Ehrlichia* organisms have evolved one or more mechanisms to circumvent the host immune system. Some bacterial pathogens are endowed with sophisticated mechanisms to adapt to a rapidly changing microenviroment in the host. One such system is the reversible switching of the expression of the array of cell surface components exposed to the host defense system.

Homologous recombination of genes in multigene families has contributed to the persistent infection of *Borrelia*

hermsii (Schwan and Hinnebusch, 1998) and *Neisseria gonorrhoeae* (Haas and Meyer, 1986). Homologous recombination of the *p28* multigenes has been hypothesized (Reddy and Streck, 1999). However, no homologous recombination of *p28* genes of *Ehrlichia* has yet been demonstrated. Homologous recombination was not observed in different passages of *E. chaffeensis* or *E. canis*, which have been passaged for several years. The DNA sequences of *p28* genes published by different laboratories are identical despite the different passage histories (Ohashi et al., 1998b, Reddy et al., 1998, Yu et al., 1999b), suggesting a lack of recombination as a mechanism of generation of genetic diversity. Moreover, the DNA sequences of five *p28* genes in a locus of *E. canis* Jake and Oklahoma isolates are identical despite the temporal and geographic separation of these isolates in nature. The genetic variation of the *p28* gene among strains of *E. chaffeensis* is very likely caused by random mutation over a long period of evolution of the gene rather than by homologous recombination.

The *p28* genes may be expressed differentially. Neither the *E. chaffeensis* nor the *E. canis p28* multigenes are one polycistronic gene. Antigenically and structurally distinct *msp-2* genes have been expressed in acute *A. marginale* rickettsemia in

experimentally infected calf (Eid et al.,1996, French et al., 1999). Protein immunoblotting detected 2-4 proteins in cell culture with a monoclonal antibody to a P28 of *E. chaffeensis* (Yu et al., 1993,1999b). Although several *E. chaffeensis p28* genes are transcribed in cell culture, a clone of tick-inoculated *E. chaffeensis* may differentially and sequentially express the *p28* multigene family *in vivo* to evade the host immune system. Different P28 proteins may have similar structure and function for *E. chaffeensis*, but different antigenicity. The hypervariable regions are predicted to contain antigenic epitopes which are surface exposed (Yu et al., 1999b). Thus, the P28s may be essential for immune escape.

It was demonstrated that only 40% of convalescent sera of monocytotropic ehrlichiosis patients had antibodies to a P28-19. Patient serum that reacted with the particular P28 of one strain of *E. chaffeensis* might not react with the protein in another strain in which the amino acid sequences of the hypervariable regions differ substantially (Chen et al., 1997, Yu et al., 1999c). The data suggest that the apparent antigenic variability of the P28 may be explained in part by differential expression of the *p28* multigene family.

The following references were cited herein:

Andrew, et al., 1989. The carrier status of sheep, cattle and African buffalo recovered from heartwater. Vet. Parasit. 34, 261-266.

Breitschwerdt, et al., 1998. Sequential evaluation of dogs naturally infected with *Ehrlichia canis*, *Ehrlichia chaffeensis*, *Ehrlichia equi*, *Ehrlichia ewingii*, or *Bartonella vinsonii*. J. Clin. Microbiol. 36, 2645-2651.

Chen, et al., 1994. Identification of the antigenic constituents of *Ehrlichia chaffeensis*. Am. J. Trop. Med. Hyg. 50, 52-58.

Chen, et al., 1997. Genetic and antigenic diversity of *Ehrlichia chaffeensis*: comparative analysis of a novel human strain from Oklahoma and previously isolated strains. J. Infect. Dis. 175, 856-863.

Dawson, et al., 1992. Susceptibility of dogs to infection with *Ehrlichia chaffeensis*, causative agent of human ehrlichiosis. Am. J. Vet. Res. 53,1322-1327.

Dawson, et al., 1994. Susceptibility of white-tailed deer (*Odocoileus virginianus*) to infection with *Ehrlichia chaffeensis*, the

etiologic agent of human ehrlichiosis. J. Clin. Microbiol. 32,2725-2728.

Dumler, J.S., Bakken, J.S., 1995. Human granulocytic ehrlichiosis in Wisconsin and Minnesota: A frequent infection with the potential for persistence. J. Infect. Dis. 173, 1027-1030.

Dumler, J.S., Sutker, W,L., Walker, D.H., 1993. Persistent infection with *Ehrlichia chaffeensis*. Clin. Infect. Dis. 17, 903-905.

Eid, et al., 1996. Expression of major surface protein 2 antigenic variants during acute *Anaplasma marginale* rickettsemia. Infect. Immun. 64, 836-841.

Haas, et al., 1986. The repertoire of silent pilus genes in *Neisseria gonorrhoeae*: evidence for gene conversion. Cell 44, 107-115.

Harrus, et al., 1998. Amplification of ehrlichial DNA from dogs 34 months after infection with *Ehrlichia canis*. J. Clin. Microbiol. 36, 73-76.

Horowitz, et al., 1998. Saddleback fever due to human granulocytic ehrlichiosis. Lancet 351, 9103.

French, et al., 1999. Emergence of *Anaplasma marginale* antigenic variants during persistent rickettsemia. Infect. Immun. 67, 5834-5840.

Ijdo, et al., 1998. Cloning of the gene encoding the 44-kilodalton antigen of the agent of human granulocytic ehrlichiosis and characterization of the humoral response. Infect. Immun. 66, 3264-3269.

Klein, et al., 1985. The detection and classification of membrane-spanning proteins. Biochim. Biophys. Acta. 815, 468-476.

McBride, et al., 1999a. Molecular characterization of a new 28-kilodalton protein gene and a multigene locus encoding five homologous 28-kilodalton immunodominant outer membrane proteins of *Ehrlichia canis*. In: Raoult, R., Brouqui, P. (Eds.), Rickettsiae and Rickettsial Diseases at the Turn of the Third Millenium. Elsevier, Paris, France, pp 43-47.

McBride, et al., 1999b. Molecular cloning of a conserved major immunoreactive 28-kilodalton protein gene of *Ehrlichia canis*: a potential serodiagnostic antigen. Clin. Diagn. Lab. Immunol. 6, 392-399.

McGeoch, D. J., 1985. On the predictive recognition of signal peptide sequences. Virus Research 3, 271-286.

Murphy, et al., 1998. Major antigenic proteins of the agent of human granulocytic ehrlichiosis are encoded by members of a multigene family. Infect. Immun. 66, 3711-3718.

Oberle, et al., 1993. Derivation of the complete *msp* 4 gene sequence of *Anaplasma marginale* without molecular cloning. Gene 136, 291-294.

Ohashi, et al., 1998a. Cloning and characterization of multigenes encoding the immunodominant 30-kilodalton major outer membrane proteins of *Ehrlichia canis* and application of the recombinant protein for serodiagnosis. J. Clin. Microbiol. 36, 2671-2680.

Ohashi, et al., 1998b. Immunodominant major outer membrane proteins of *Ehrlichia chaffeensis* are encoded by a polymorphic multigene family. Infect. Immun. 66, 132-139.

Palmer, et al., 1998. Persistence of *Anaplasma ovis* infection and conservation of the *msp-2* and *msp-3* multigene families within the genus *Anaplasma*. Infect. Immun. 66, 6035-6039.

Palmer, et al., 1994. The immunoprotective *Anaplasma marginale* major surface protein 2 is encoded by a polymorphic multigene family. Infect. Immun. 62, 3808-3816.

Reddy, et al., 1998. Molecular characterization of a 28 kDa surface antigen gene family of the tribe *Ehrlichieae*. Biochem. Biophys. Res. Commun. 247, 636-643.

Reddy, G.R., Streck, C.P., 1999. Variability in the 28-kDa surface antigen protein multigene locus of isolates of the emerging disease agent *Ehrlichia chaffeensis* suggests that it plays a role in immune evasion. Mol. Cell Biol. Res. Commun. 1, 167-175.

Rikihisa Y., 1991. The tribe *Ehrlichieae* and ehrlichial diseases. Clin. Microbiol. Rev. 4, 286-308.

Roland, et al., 1995. Ehrlichiosis - A cause of prolonged fever. Clin. Infect. Dis. 20, 821-825.

Schwan, et al., 1998. Bloodstream-versus tick-associated variants of relapsing fever bacterium. Science 280, 1938-1940.

Sulsona, et al., 1999. The *map1* gene of *Cowdria ruminantium* is a member of a multigene family containing both conserved and variable genes. Biochem. Biophys. Res. Commun. 257, 300-305.

Swift, B.L., Thomas, G.M., 1983. Bovine anaplasmosis: elimination of the carrier state with injectable long-acting oxytracycline. JAMA 183, 63-65.

Telford, et al., 1996. Perpetuation of the agent of human granulocytic ehrlichiosis in a deer tick-rodent cycle. Proc. Natl. Acad. Sci. USA 93, 6209-6214.

Von Heijne G., 1986. A new method for predicting signal sequence cleavage sites. Nucl. Acids Res. 14, 4683-4690.

Van Vliet, et al., 1994. Molecular cloning, sequencing analysis, and expression of the gene encoding the immunodominant 32-kilodalton protein of *Cowdria ruminantium*. Infect. Immun. 62, 1451-1456.

Weiss, et al., 1989. Energy metabolism of monocytic *Ehrlichia*. Proc. Natl Acad. Sci. USA. 86,1674-1678.

Yu, X.J., Brouqui, P., Dumler, J.S., Raoult, D., 1993. Detection of *Ehrlichia chaffeensis* in human tissue by using a species-specific monoclonal antibody. J. Clin. Microbiol. 31, 3284-3288.

Yu, et al., 1999a. Characterization of the genus-common outer membrane proteins in *Ehrlichia*. In: Raoult, R., Brouqui P, (Eds.), Rickettsiae and Rickettsial Diseases at the Turn of the Third Millenium. Elsevier, Pris, France, pp 103-107.

Yu, et al., 1999b. Genetic diversity of the 28-kilodalton outer membrane protein of human isolates of *Ehrlichia chaffeensis*. J. Clin. Microbiol. 37, 1137-1143.

Yu, et al., 1999c. Comparison of *Ehrlichia chaffeensis* recombinant proteins for diagnosis of human monocytotropic ehrlichiosis. J. Clin. Microbiol. 37, 2568–2575.

Zaugg, et al., 1986. Transmission of *Anaplasma marginale* Theiler by males of *Dermacentor andersoni* Stiles fed on an Idaho field-infected, chronic carrier cow. Am. J. Vet. Res. 47,2269-2271.

Any patents or publications mentioned in this specification are indicative of the levels of those skilled in the art to which the invention pertains. These patents and publications are herein incorporated by reference to the same extent as if each individual publication was individually indicated to be incorporated by reference.

One skilled in the art will readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The present examples along with the methods, procedures, treatments, molecules, and specific compounds described herein are

presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention as defined by the scope of the claims.